(19) Europäisches Patentamt
Eur pean Patent Office
Office européen des brevets

(11) Publication number: **0 187 289**
A2

(12) # EUROPEAN PATENT APPLICATION

(21) Application number: **85115711.5**

(51) Int. Cl.4: **G06F 12/08**

(22) Date of filing: **10.12.85**

(30) Priority: **31.12.84 US 687807**

(43) Date of publication of application:
16.07.86 Bulletin 86/29

(84) Designated Contracting States:
DE FR GB NL SE

(71) Applicant: **International Business Machines Corporation**
Old Orchard Road
Armonk, N.Y. 10504(US)

(72) Inventor: **Pakulski, Francis Joseph**
R.D. No. 1, Spear Street Box 1894
Shelburne Vermont 05482(US)

(74) Representative: **Teufel, Fritz, Dipl.-Phys. et al**
IBM Deutschland GmbH. Europäische Patentdienste
Postfach 265
D-8000 München 22(DE)

(54) **Hierarchical memory system.**

(57) A hierarchical memory system in which a lower level (40) transfers data serially to an upper level (12) and in parallel to a yet lower level. The lower level includes a two-port memory chip (25) having a wide buffer (28) for parallel accesses to the memory array, to the yet lower level, and to a serial buffer (32). The serial buffer is serially accessed by the upper level simultaneously with accesses to the wide buffer.

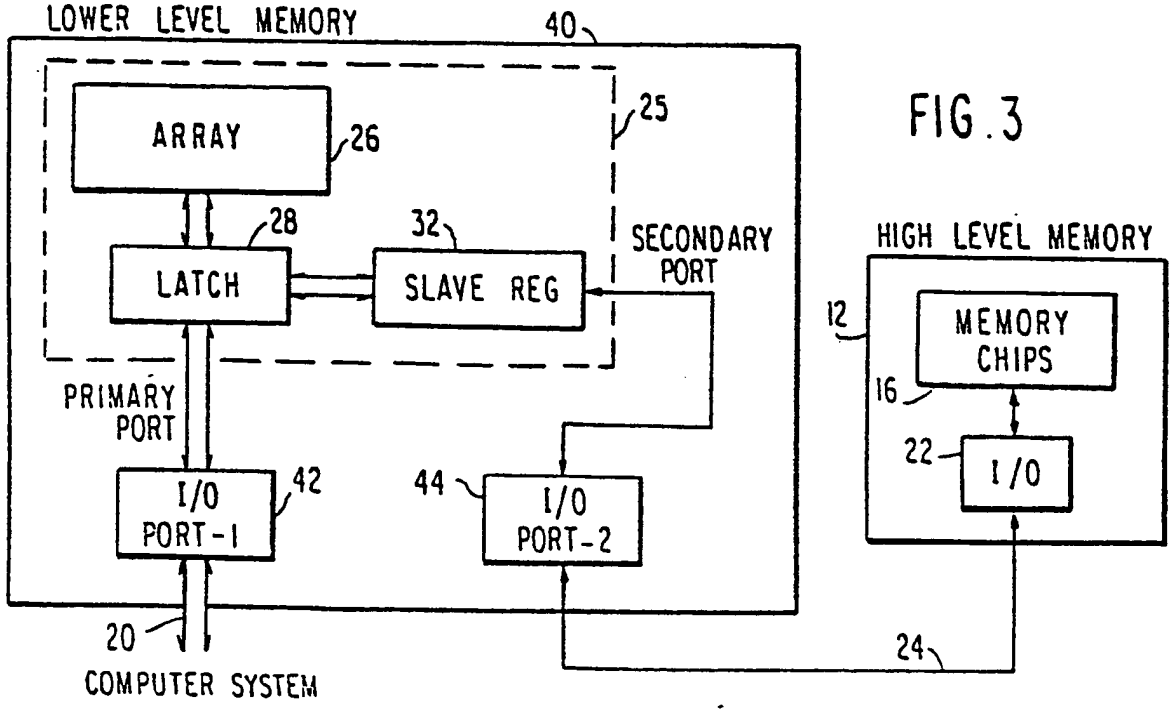


FIG. 3

EP 0 187 289 A2

HIERARCHICAL MEMORY SYSTEM

This invention relates to a hierarchical memory system in accordance with the preamble of claim 1.

Almost all computer systems include some type of memory. Th predominant type of memory used for computer application is random access memory (RAM) although larger systems requiring vast amounts of memory storage typically include additional serial memory, usually in the form of magnetic tapes or discs. In simpler computer systems, the one or more memories are directly connected to the processor which then directly reads into and writes from a selected memory device.

However, there is a trend toward hierarchical memories in which a larger, higher level memory is not connected directly to the processor but instead transmits to and receives data from a lower level memory. The lower level memory receives data from and transmits data to, not only the higher level memory, but also the computer. A cache memory is one example of a lower level memory. The concept of multi-level memories may be extended to more than two levels of memory.

An example of a hierarchical memory is illustrated in Figure 1, in which two memory cards 10 and 12 contain respectively a low level memory and a high level memory. Each card 10 or 12 contains an array of memory chips 14 and 16. Typically, the memory chips of the low level memory 10 are faster but of fewer number than the memory chips 16 of the high level memory 12. Thus the high level memory 12 is used for bulk storage while the low level memory 10 is used for its fast access and buffering capability. The low level memory 10 provides data from its memory chips 14 to the computer systems through a primary port 18 and a fast data channel 20. Similarly, the low level memory provides data to the high level memory 12 through a secondary port 21 which is connected to an I/O port 22 of the high level memory 12. The secondary port 21 and the I/O port 22 are connected by an inter-level channel 24. Data transmission on the two channels 20 and 24 can be done in either direction. In typical designs, the low level memory l0 can operate only one of its ports 18 and 21 at any one time. That is, the low level memory 10 either selects the primary port 18 for access by the data channel 20 to the computer system or alternatively selects the secondary port 21 for access by the inter-level channel 24 and the high level memory 12. The data channel 20 usually is a high speed channel, matched to the computer system, and is a parallel bus. In contrast, the inter-level channel 24 has a different channel capacity. The low level memory 10 provides buffering between the differing data rates on the two channels 20 and 24.

A disadvantage of the memory system of Figure 1 is the lack of simultaneous access of the two ports 18 and 21 to the memory chips 14. While data is being accessed through the secondary port 21, either in a read or a write mode the memory chips 14 cannot be accessed by the data channel, thus impacting the operation of the fast paced computer system. Another disadvantage is that during the serial accesses through the secondary port 21, the memory chips 14 and their support circuitry need to be fully powered at a power level dictated by the inherent cycle time of the chip. As a result, the memory system of Figure 1 is both slow and demands full power for each access.

The use of dual ported memory chips in a hierarchical memory system is disclosed in US-patent 4.489.381 and EP-A-85107253.8.

Accordingly it is the object of the present invention to provide a hierarchical memory system of the aforementioned kind that allows simultaneous multiple access to a lower level memory.

This object is achieved by the invention of claim 1; embodiments of the invention are characterized in the dependent claims.

The invention proposes a hierarchical memory system in which a lower level memory is built from two-port memory chips. Data is accessed randomly from the memory chip to a primary buffer. Data can be transferred between the primary buffer and a secondary buffer within the memory chip. The primary buffer and the secondary buffer can be independently and simultaneously accessed from outside the chip at differing data rates. A parallel access to the primary buffer is used for accesses to the computer system or a yet lower memory level. Accesses to the secondary buffer are performed serially and are made available to a higher level memory. The memory chip does not require full power for each access.

An embodiment of the invention is now described in detail with reference to the accompanying drawings, in which:

Figure 1 is a block diagram of a hierarchical memory system in the prior art;

Figure 2 is a schematic diagram of a two-port memory chip usable with the present invention; and

Figure 3 is a block diagram of one embodiment of the hierarchical memory of the present invention.

The present invention uses a two-port memory chip in its hierarchical memory. Several varieties of multi-port chips have recently been developed. For instance, Nordling et al. in U.S. Patent 4,410,964 discloses a memory device with two ports, each with multiple bits. Rao in U.S. Patent 4,347,587 discloses a memory chip with both parallel and serial parts. However, in this chip, the serial port is associated only with a portion of the memory dedicated to the serial port. Ackland et al. in U.S. Patent 4,412,313 discloses a two-port chip in which a shift register is connected to the parallel output lines and can provide a high speed serial output. Baltzer in U.S. Patent 4,150,364 discloses a somewhat different memory system in which two memory chips can be simultaneously accessed.

A two-port memory chip 25 particularly useful for the present invention is illustrated in Figure 2. This type of chip is described by R. Matick et al. in a technical article entitled "All Point Addressable Raster Display Memory" appearing in the IBM Journal of Research and Development, Vol. 28, No. 4, July 1984 at pp. 379-392. Data is stored in a memory array 26 in long words of 128 bits. Each 128-bit word can be randomly accessed by a row address. The individual words are accessed through sense amplifiers and latches 28 which thus serve as a buffer to the memory array 26. The 128-bit word in the latches 28 is further subdivided into smaller bytes, typically of from 2 to 4 bits apiece. Each byte in the latches 28 is randomly accessed by a column address for a transfer to or from the 4-bit wide primary port. Data in 128-bit words can also be transferred in parallel between the latches 28 and a master register 30. In order to decouple the transfer word from the latches 28, the master register 30 can be accessed by a slave register

32 by the parallel transfer of the 128-bit word. Selection circuits control the slave register 32 so that it can be accessed serially through a secondary port. All the previously described transfers can occur in either direction.

In the two-port memory chip 25, the slave register 32 can be accessed through the secondary port independently of the access of the latches 28 through the primary port. For instance, a word can be read from the array 26 into the latches 28 and then transferred through the master register 30 to the slave register 32. Thereafter, this word can be read serially from the slave register 32 out through the secondary port. Simultaneously with the serial access through the secondary port, the latches 28 can be accessed by the primary port. Additional words can be read from the array 26 into the latches 28 and then outputted through the primary port, all during the reading of the single word in the slave register 32 out of the secondary port. Furthermore, multiple words can be read into the array 26 from the primary port while the secondary port is operating in the write direction. For a fairly high speed chip, the parallel access through the primary port may be operating at a cycle rate of 80-150 ns per word while the serial access through the secondary port is not likely to exceed a cycle rate of 20 ns per bit. For 128-bit words, the secondary port is thus cycling at 1.56 ms per word, thus permitting many accesses of data through the primary port in the same time that one complete word is accessed through the secondary port.

One embodiment of the present invention uses the two-port memory chip 25 in a hierarchical memory system illustrated in Figure 3. Similar elements to those of Figure 2 are numbered with the same reference numerals and will not be further discussed. The two- port memory chip 25 is included in a lower level memory card 40. The master register 30 and the address and selection lines are not explicitly shown in Figure 3. The primary port of the memory chip 25 is connected through a first I/0 port 42 to the data channel 20 to the computer system. This I/0 port 42 could be up to 4-bits wide in present technology for the described memory chip 24. The secondary port of the memory chip 24 is connected through a second I/0 port 44 to the inter-level channel 24 to the high level memory 12.

It is seen that the computer system can simultaneously access the latch 28 while the high level memory 12 is accessing the slave register 32. Thus the slave register 32 can be accessed relatively slowly over the serial inter-level channel 24 for one word while several words can be accessed between the array 26 and the computer system through the latch 28. The accesses can be any combination of reading and writing. Of course, when the access to the slave register 32 has been completed, a subsequent access to the slave register 32 would require either a read from or a write to the array 26. This subsequent access necessarily involves the latch 28 so that the the accessing through the primary port to the computer system is necessarily interrupted. However, the duration of this interruption of the primary port is only a small fraction of the total access time through the secondary port. The simultaneous operation of the primary and the secondary ports eliminates significant waiting time caused by data transfers between levels and can lead to significant improvements in overall performance of the hierarchical memory system.

Compared to the hierarchical memory system of Figure 1, the present invention offers several advantages. The improved performance by a simultaneous access to the two ports provides improved performance. During those periods when only the secondary port is being accessed and no data is being transferred between the slave register 32 and the array 26, the I/0 port 42 associated with the primary port and the sense amplifiers and latches 28 can be depowered. This depowering reduces both the average and the peak power levels required for the lower level memory card 40. This reduced power results in a lower operating temperature than for the conventional approach. The reduction in temperature is expected to be about 10°C. Because of the depowering, the lower level memory 40 is expected to have improved reliability, predicted to be about a 20% reduction in cumulative failures. The use of serial memory in the memory chip 16 of the higher level memory 12 provides a cost reduction because a serial access memory can be made at lower cost than RAM.

Although the lower level memory 40 of Figure 3 has been described as consisting of a single memory chip 25, the invention can be applied to a lower level memory 40 of multiple memory chips, each having two ports connected respectively to the I/O parts 42 and 44. The different memory chips can either be separately selected depending upon the address or can be operating in parallel to provide different bits of an addressed word.

It should be noted that the memory organization for the lower level memory 40 of Figure 3 is not appropriate if the slave register 32 is not on the same memory chip as the memory array 26. Such an off-chip slave register would require an inordinate number of I/0 pins from the latch 40 to the off-chip slave register. Furthermore the simultaneous transfer of so many bits of data across the chip boundary would create a power surge if it were performed between different chips.

The described embodiment assumes a control transfer of data between the latch 28 and the slave register 32. However, in some applications where the flow of data is predictable, any access of array 26 through the primary port could be designated to cause an automatic transfer of data with the slave register 32 for a corresponding access through the secondary port.

Claims

1. A hierarchical memory system using dual-port memory chips (25) in at least one hierarchy level characterized in that there are provided:

- first buffer memories (28) in the dual-port chip being connected to the array (26) and the first port of the chip;

- second buffer memories (32) in the dual-port chip being connected to the first buffer memory (28) and the second port of the chip;

- means (42) for randomly accessing the first buffer memory (28) from the next lower hierarchy level;

- means (44) for serially transferring information between the second buffer memory (32) and the next higher hierarchy level.

2. Memory system in accordance with claim 1, characterized in that

a complete memory row is transferred in each access between the first buffer m mory and the memory array.

3. Memory system in accordance with claim 1 or 2, char-

acterized in that

the information exchange between the first and second buffer memories is effected in complete memory rows.

4. Memory system in accordance with one of the claims 1 to 3,

characterized in that

a third buffer memory (30) is interconnected between the first buffer memory (28) and the second buffer memory (32) on the dual-port memory chip.

5. Memory system in accordance of one of the claims 1 to 4,

characterized in that

means are provided to depower the first buffer m mory (28) if only accesses to the second buffer memory are effected.

6. Memory system in accordance of one of the claims 1 to 5,

characterized in that

means are provided to transfer the contents of the first buffer memory to the second buffer memory upon each access of the memory array.

FIG.1

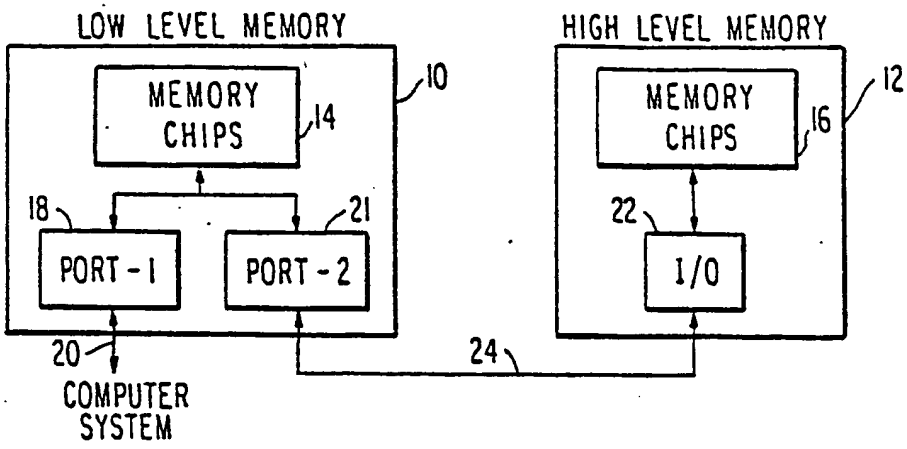


FIG.2

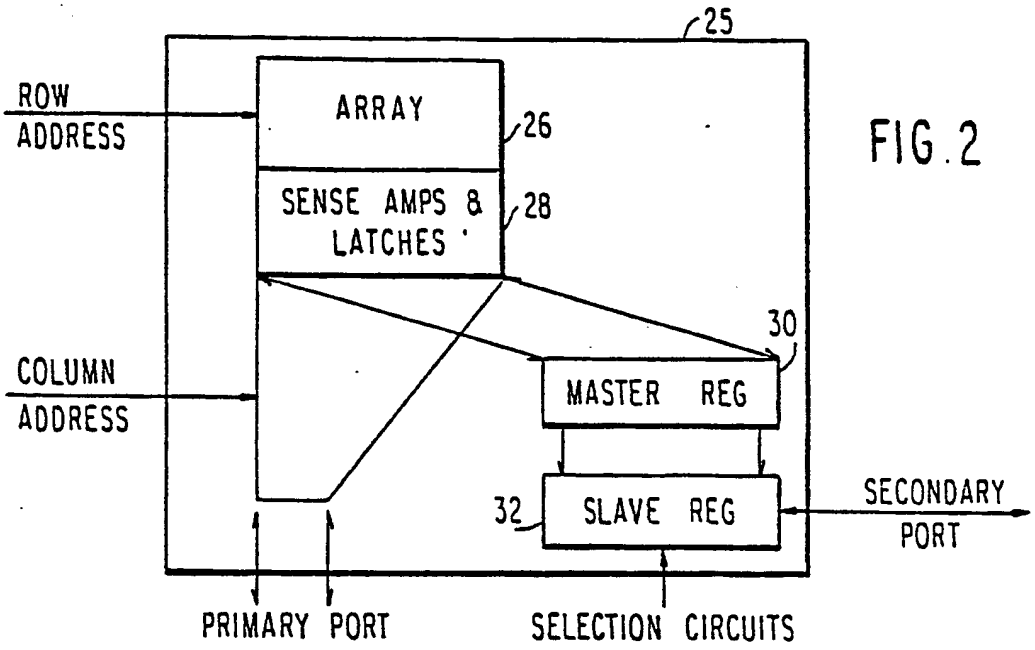
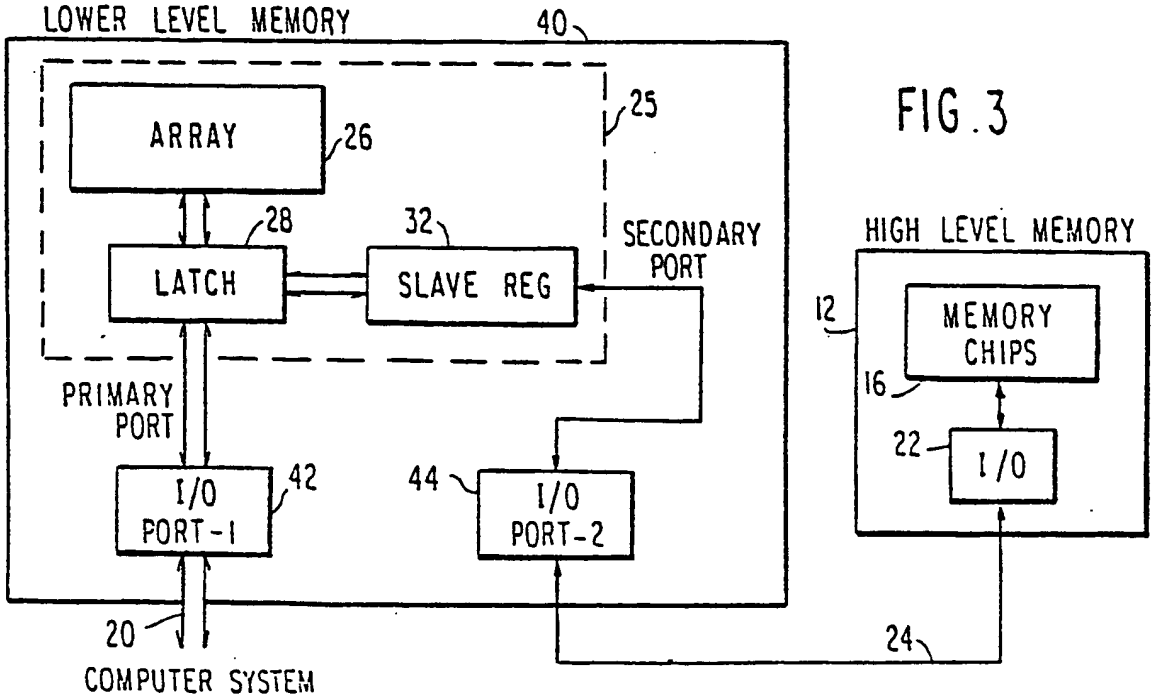


FIG.3



(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 187 289**
A3

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **85115711.5**

(51) Int. Cl.⁴: **G06F 12/08**

(22) Date of filing: **10.12.85**

(30) Priority: **31.12.84 US 687807**

(43) Date of publication of application:
16.07.86 Bulletin 86/29

(84) Designated Contracting States:
DE FR GB NL SE

(88) Date of deferred publication of the search report:
03.05.89 Bulletin 89/18

(71) Applicant: **International Business Machines Corporation**
Old Orchard Road
Armonk, N.Y. 10504(US)

(72) Inventor: **Pakulski, Francis Joseph**
R.D. No. 1, Spear Street Box 1894
Shelburne Vermont 05482(US)

(74) Representative: **Barth, Carl Otto et al**
IBM Deutschland GmbH Patentwesen und Urheberrecht Schönaicher Strasse 220
D-7030 Böblingen(DE)

(54) Hierarchical memory system.

(57) A hierarchical memory system in which a lower level (40) transfers data serially to an upper level (12) and in parallel to a yet lower level. The lower level includes a two-port memory chip (25) having a wide buffer (28) for parallel accesses to the memory array, to the yet lower level, and to a serial buffer (32). The serial buffer is serially accessed by the upper level simultaneously with accesses to the wide buffer.

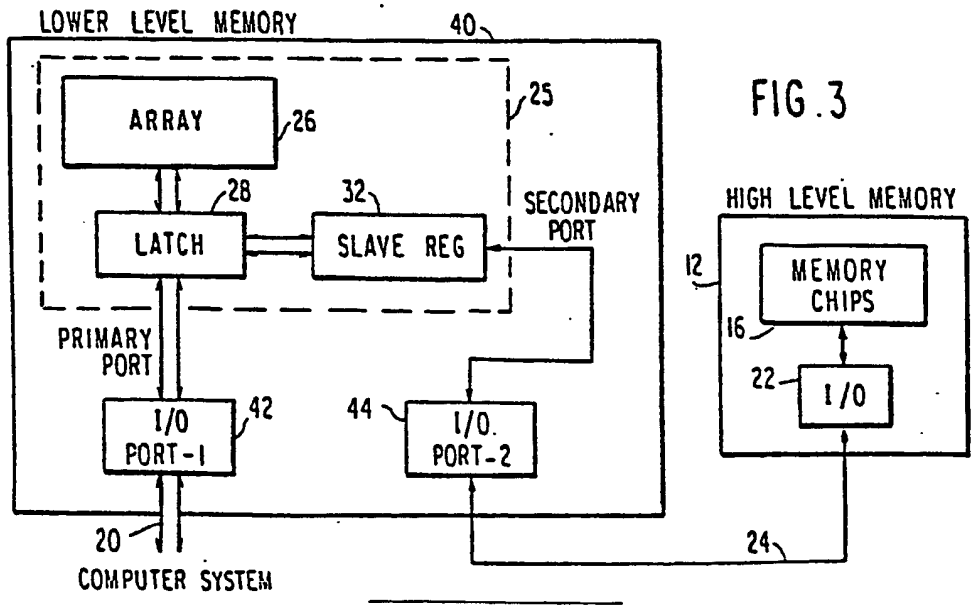


FIG.3

EP 0 187 289 A3

European Patent Office

EUROPEAN SEARCH REPORT

Application Number

EP 85 11 5711

DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document with indication, where appropriate, of relevant passages	Relevant to claim	CLASSIFICATION OF THE APPLICATION (Int. Cl.4)
D,Y	EP-A-0 100 943 (IBM) * Page 2, line 30 - page 3, line 24; figure *	1-4	G 06 F 12/08
Y	EP-A-0 097 778 (IBM) * Page 6, line 14 - page 9, line 6; page I1, line 25 - page 15, line 4; figures 1,5 *	1-4	
A		6	
A,D	IBM J. RES. DEVELOP, vol. 28, no. 4, 4th July 1984, pages 379-392; R. MATICK et al.: "All points addressable raster display memory" * Figure 7; page 385, lines 4-11 *	1	
A	PATENT ABSTRACTS OF JAPAN, vol. 5, no. 165 (P-85)[837], 22nd October 1981; & JP-A-56 93 177 (NIPPON DENKI K.K.) 28-07-1981	5	

TECHNICAL FIELDS SEARCHED (Int. Cl.4)

G 06 F 12/08
G 11 C 7/00
G 11 C 8/00

The present search report has been drawn up for all claims

Place of search	Date of completion of the search	Examiner
THE HAGUE	07-02-1989	MASCHE C.M.

CATEGORY OF CITED DOCUMENTS

X : particularly relevant if taken alone
Y : particularly relevant if combined with another document of the same category
A : technological background
O : non-written disclosure
P : intermediate document

T : theory or principle underlying the invention
E : earlier patent document, but published on, or after the filing date
D : document cited in the application
L : document cited for other reasons

& : member of the same patent family, corresponding document

EPO FORM 1503 03.82 (P0401)